DOCUMENT RESUME

ED 257 878 TM 850 347

AUTHOR Brown, William J., Jr.
TITLE Challenges in Implementing a Statewide Testing Program.
PUB DATE Mar 85
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Disabilities; Elementary Secondary Education; Minimum Competency Testing; *Program Implementation; Standardized Tests; *State Programs; Statewide Planning; *Testing Programs; Test Selection
IDENTIFIERS *North Carolina

ABSTRACT

Several of the major considerations involved in implementing a statewide testing program are discussed. The test results provide information which influences the educational enterprise in matters of curriculum, instruction, guidance, and accountability. Tests to be used in an assessment program must be developed by the administering agency when commerially published tests are inappropriate. North Carolina administers the same test battery statewide. This requires uniformity in basic aspects of the curriculum. The problems inherent in test selection for the program are discussed, as well as the challenges of testing special populations. Testing modifications are made for blind and deaf students, as well as those students with other handicapping conditions. States face the challenge of creating a positive interface between students and teachers to correlate instruction with testing. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED257878

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CHALLENGES IN IMPLEMENTING A STATEWIDE TESTING PROGRAM

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
W.J. Brown, Jr.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Prepared for Presentation at

1985 AERA Annual Meeting

by William J. Brown, Jr.

Special Assistant for Research

North Carolina Department of Public Education

Testing programs which are uniformly administered statewide represent efforts of scale which are massive, require logistics which simulate a well run army, and have policy implications of the highest order. In our state, over a half million students are tested annually; each receiving a test booklet, answer sheet, and instructions for the testing period. About six weeks later, properly scored test reports are returned for students throughout the state with national, state, regional, school system, and local school comparisons. From these data, students are evaluated for special needs, instruction is planned, curriculum improvement is designed, parents are conferenced, and the public is informed through media releases. In a very real sense, tests provide information which influences the educational enterprise in matters of curriculum, instruction, guidance, and accountability. Certainly instruments and programs of such potential potency deserve the utmost attention if they are to be constructive forces in education. The balance of this presentation will deal with several of the more crucial considerations in implementing a statewide testing program.

Test Selection

Test selection from among several commercially available products is one obvious alternative yet it does not exclude the possibility that tests will need to be developed for special purpose testing programs. In our case we have done some of both. General achievement batteries in reading, language, and mathematics were found to be adequate through grade nine. Proper curriculum match was more difficult to achieve for other subjects, especially social studies; so test development has become necessary. Test development is nearly always required for special purposes such as minimum competency testing or for the testing of specific courses such as Algebra I or Biology.

In all cases, the tests have to be synchronized with the purposes to be served; and the primary purpose to serve in our state is the improvement of curriculum and instruction. Since tests are administered statewide in North Carolina, it is clear that uniformity in basic aspects of the curriculum also

TM 850 347

is a statewide expectation. This precept parallels our adoption of texts statewide, our standard course of study, our expectation that every curriculum area be included in the accreditation plans for each school system, and more recently a movement to define and seek state funding for a basic education program. Against this backdrop, it is easily seen why test information and test selection is so important to us.

In selecting tests for our programs, we have set forth an elaborate review process which uses teachers representing appropriate grades and subject areas, subject matter specialists from local and state staffs, and technical staff who judge the adequacy of test development processes. These review groups work intensively both in groups and independently and they have direct access to publisher resource staff as necessary. Judgements indicate which tests are appropriate to the needs or sometimes that no tests are adequate. In rendering these judgements of format, content, and technical adequacy, each test item is evaluated against highly specified criteria and then the test as a whole is examined. The final result is the selection of a test which may not be the exact product that we desired but it will be suitable to provide information for the various purposes that must be accommodated.

In situations where test development is required, we often find that our expectations must be tempered as we balance between the needs for accountability of programs, measurement of overall student achievement, and measurement of instructional categories. You might say that after building a test, one has a greater appreciation of why no tests are evaluated as perfect!

The conflicting purposes for testing deserves more attention since it comes into play when selecting or developing tests. On one hand, there are expectations for normative comparisons to be used for student guidance and for system accountability. An equally strong interest calls for the provision of test information that is related to instruction. To balance these interests clearly requires compromise. The length of testing time must be reasonable and therefore the measurement of specific objectives must either be with fewer items than desirable or the number of objectives must be reduced. The statistical characteristics of test items also can be quite different for a norm referenced test where a P Value of .50 is desirable than for a mastery level item where higher achievement is expected. The balancing act continues as scoring and reporting is considered. Are derived scores such as scale scores, grade equivalents, or percentiles sufficient? Clearly the answer is no if instructional purposes are important. Teachers want to know which skills have been mastered. To the credit of test publishers, considerable effort in research and experimentation has produced a system of expanded reporting beyond normative scores. The challenge for users of tests is to determine how much instructional information can be reliably reported and what interpretive cautions must be provided to teachers and curriculum developers as they prepare to use the information.

I expect the dual demand for accountability and diagnostic information to continue. However, as we become more sophisticated in our use of technology, more of the diagnostic information will flow from teacher administered and scored tests which are relevant to the curriculum for that class. These tests may be readily available from item banks which are correlated to textbooks and other instructional resources and may or may not be provided by commercial sources. When instructional testing becomes a function managed by the

classroom teacher, the normed referenced tests will be used to yield comparative information which it does quite well and also provide information which will assist the teacher is selecting appropriately leveled diagnostic tests.

Testing Special Populations

The next area of challenge has to be which groups or populations are to be tested, how are they are to be tested, and how can this be accomplished in the same manner throughout the state. For tests with national norms, it would be instructive to know which groups the publishers have been included in their norms and in what proportions. Otherwise, the states will find it impossible to fairly administer the test scores across each local school system. For most tests developed prior to 1984, the testing provisions for special populations were explicitly ambiguous! That is, you knew what the instructions were, but there were no data on how these instructions were implemented by local school systems during the standardization of the test. In fact, there is considerable speculation that the localities varied widely in their practices thereby insuring that a common protocol for a statewide administration would invalidly represent the national standardization process and its resulting comparisons. Such problems as this are expressly distressful to statewide testing administrators because they must standardize the process in their states. And, if standardization does not occur statewide, the generalizations and comparisons across schools and school systems are inappropriate and invalid. In our state, we had to establish a protocol which stood statewide; yet the common protocol may have invalidated our national comparisons somewhat and probably resulted in more handicapped students being tested than was the case nationally. Hopefully, the national standardization processes is use this year will be better.

In minimum competency testing (MCT) the comparison of groups or systems is not the problem of testing special populations. In MCT the problem is one of fairness to the student and of insuring equity among all students. Take the equity issue first. If a minimum score on the MCT entitles a student to earn a diploma, for example; then students are not being treated equitably unless that score is required of all students seeking the diploma. The fact that some may not reach the score ultimately and thus may not be able to achieve a diploma is an expected result of a set standard. Now the question shifts from who to test to how may all students be fairly tested. Usually, we answer this question by accommodating the handicapped as the test is administered; but requiring the same test content and standard.

Testing modifications which are used in our state include the obvious modifications for the blind (braille or large print editions) and for the deaf (sign language or video tapes); however other handicapping conditions require special testing conditions if those students are to be fairly tested. We allow for testing over an extended time period to avoid fatigue and lessen frustration. These sessions are scheduled by the student's teacher for an appropriate testing period.

If a student is physically or emotionally unable to mark answers on an answer sheet, proctors can perform this task. Also, some students function better if they care allowed to mark the correct answer directly in the test

booklet. Another modification of the test environment is to allow students to be tested in small groups usually by their regular teacher.

In rare cases we find it necessary to test learning disability students who are primarily oral learners with an audio-cassette edition which is used in conjunction with the printed test.

Usually a handicapped child will be tested with only one modification, but it is permissable to authorize multiple modifications if needed.

We feel that the testing of handicapped children with modifications has allowed them to perform up to their level of capability and has also maintained the standard set forth by The State Board of Education. The input that has been received during the seven years that our MCT program has been operational reveals that parents and advocacy groups for handicapped children consider the modifications as essential aspects of the program.

Gate-Keeping Tests

Recently state officials have been expected to establish achievement standards which required a set level of achievement to be attained prior to advancing to the next grade or before a diploma was issued. In some cases, additional funding was provided for those students who measured below the mark. This was the case in our state as we initiated MCT for high school graduation and without qualification I can report that it has helped students to attain those basic skills that are necessary to function as an adult. Because of these standards, students can interpret warranties, handle financial matters, use transportation information, read and interpret safety information and do countless other tasks which are expected of an individual that is self-sufficient.

Now we are expected to apply this concept at grades 3, 6, and 8 for reading, language, and mathematics. Rather than describe the details of the plan, let me indicate that the testing will be efficient in determining who is deficient in skills, it will be completely diagnostic on which skills are weak, and there will be resources available to extending these students time for learning and to enhance all the instruction which the student receives.

Testing for retention alone is considered as negative gate-keeping. It provides only to exclude students from educational attainment. It almost certainly would increase the percentages of students who fall behind and then drop out of the educational process entirely. In our state, we must see that this does not happen. We will test for promotion and we will retain some students. However, a retained student will also receive about 33% more state funding to provide for remediation. We feel that these students will achieve appreciably higher with gate-keeping and remediation and that improvements in the student's cognitive skills will enable them to profit more from their regular instruction. The approaches used in the remedial instruction are expected to influence how each child is taught during the next year. If low achievement is partially caused by a mismatch in learning style and teaching method, (and I think it is) we may be making a major instructional change for students who are presently in the lower quartile. If so, the testing for standards and the anxiety which it causes will have been worthwhile for students.

Epilogue

Let me conclude with the following thought. If instruction proceeds without systematic knowledge about the student's capabilities, we can expect students to do less than their potential. If students and teachers can benefit from correlating instruction with testing and testing with instruction, why is there such resistance to the practice? Perhaps it is because neither know the other until they are forced by circumstances to meet. That cirumstance in many states has been legislated statewide testing programs. Our challenge is to construct a positive interface between the two.